Biosecurity in the age of Big Data: a conversation with the FBI

Keith G. Kozminski
Department of Biology, University of Virginia, Charlottesville, VA 22904; Department of Cell Biology, University of Virginia, Charlottesville, VA 22908

ABSTRACT **New scientific frontiers and emerging technologies within the life sciences pose many global challenges to society. Big Data is a premier example, especially with respect to individual, national, and international security. Here a Special Agent of the Federal Bureau of Investigation discusses the security implications of Big Data and the need for security in the life sciences.**

Monitoring Editor
David G. Drubin
University of California, Berkeley

Received: Jul 30, 2015
Accepted: Aug 5, 2015

INTRODUCTION

"The FBI is reading our poster!" Granted, this is not a typical refrain heard at the annual meetings of the American Society for Cell Biology, but it is heard frequently at other research meetings, for example, in the field of synthetic biology. I admit that it was head turning when I first heard these words spoken a few years ago at the International Genetically Engineered Machines (iGEM) Jamboree, which is an annual, global, intercollegiate synthetic biology competition. In format, the iGEM Jamboree is much like the annual meeting of any major scientific society. But why, in the aisles, were there suited people with badges? Perhaps a new age has dawned upon the research community. The contents of this special issue of *Molecular Biology of the Cell*, with an emphasis on Big Data, certainly suggest that this is true. Nonetheless, overt governmental examination of research beyond the standard purview of granting agencies and its program officers can only raise questions. To answer some of these questions, I invited, on behalf of *Molecular Biology of the Cell*, Supervisory Special Agent (SSA) Edward You, who heads the Biological Countermeasures Unit (BCU) at Federal Bureau of Investigation (FBI) Headquarters in Washington, D.C., and frequently addresses the synthetic biology community, to have a conversation on biosecurity, especially with respect to Big Data (Figure 1). This conversation was recorded on July 17, 2015, and is presented here, abridged and edited for clarity and considerations of space.

ONE FOOT IN NATIONAL SECURITY; ONE FOOT IN THE LIFE SCIENCES

MBoC: Agent You, before you talk about Big Data, please tell our readers about your scientific background and path to the FBI.

SSA You: I got my bachelor's degree in the biological sciences from the University of California at Irvine, then a master's degree in biochemistry and molecular biology at the University of Southern California. All that has served me well; it does show that there is life without a PhD.

Before joining the Bureau, I came from the laboratory setting. I had six years of graduate research in human gene therapy, with a focus on retrovirology, and three years in the biotech sector at Amgen, where I did oncology research. Then I decided to go into public service and apply to the FBI.

MBoC: What are your responsibilities at the FBI? What is your mission today?

SSA You: I sit at headquarters at the Weapons of Mass Destruction (WMD) Directorate in the Biological Countermeasures Unit. The WMD Directorate is one of the newest divisions of the FBI. It was born out of the events of September 11, 2001. On the heels of that terrorist event, we had the anthrax mailings. It was a serious wake-up call for the U.S. government and the FBI in particular. Since then, as a law enforcement service, our priority has become one of prevention rather than being reactive, just going in and investigating a crime or incident. Now our number one priority is to prevent in particular a 9/11 from happening again. Safeguarding science is the theme of my mission. Part of that is reaching out proactively to different members of the scientific community, ranging from the private sector, biotech and the pharmaceutical industry;

DOI:10.1091/mbc.E14-01-0027

Address correspondence to: Keith G. Kozminski (kkoz@virginia.edu).

Abbreviations used: AAAS, American Association for the Advancement of Science; BCU, Biological Countermeasures Unit; BWC, Biological Weapons Convention; CDC, Centers for Disease Control and Prevention; DoD, Department of Defense; FBI, Federal Bureau of Investigation; iGEM, International Genetically Engineered Machines; NIH, National Institutes of Health; SSA, Supervisory Special Agent; WMD, weapons of mass destruction.

FIGURE 1: FBI SSA Edward You. In addition to heading the FBI's BCU, he is a Working Group member of the National Security Council Interagency Policy Committee on Countering Biological Threats and an *ex officio* member of the NIH National Science Advisory Board for Biosecurity. He also serves on two National Academies committees: the Institute of Medicine's Forum on Microbial Threats and the Committee on Science, Technology, and Law's Forum on Synthetic Biology. SSA You also serves on the Strategic Advisory Board for the Synthetic Biology and Engineering Research Center and as an instructor for the United Nations Interregional Crime and Justice Research Institute. He can be reached at (202) 324-0236 or Edward.You@ic.fbi.gov.

to universities; to the iGEM; and even to the amateur community, the sprawling Do-It-Yourself bio community, showing how members of law enforcement and the life science community have a shared responsibility of safeguarding the development and very beneficial applications of the life sciences. I find myself in a unique position, where I have one foot in national security and another in the life sciences. I seek very hard to ensure that we are able to support both at the same time.

BIG DATA WORRIES AT THE FBI

MBoC: You mentioned synthetic biology and have been involved in that community. However, it seems more recently that the FBI has been showing more overt concern toward the security of Big Data in the life sciences. Why does the FBI have concern?

SSA You: If you take my consideration of how to protect the life sciences in a proactive manner, it is our responsibility to identify emerging areas. Six years ago the emerging area was synthetic biology. That is why you have seen all this activity and outreach occurring, especially at iGEM.

The reason why Big Data has become very significant is that it is the next evolutionary step that synthetic biology will take, meaning that all applications and technologies coming out of this field will be completely dependent upon data—all the various omics.

A very good example is precision personalized medicine, where you are seeing tremendous investments in drug development, particularly in cancer research and metabolic disease, where very large data sets are leveraged. If you are looking at an individual's genome, it is just one snap shot. What are needed are data over time, during exposure to the environment, for example. From the human standpoint, maybe this is looking at your lifestyle—daily diet or exercise. It all goes into helping determine potential health vulnerabilities and appropriate therapies. If you set that as a stage and then look at potential policy aspects, there is a lot of activity looking at privacy, but not a whole lot looking specifically at security.

So, back in April 2014, I partnered with the American Association for the Advancement of Science (AAAS) and the United Nations Interregional Crime and Justice Research Institute (UNICRI). We kicked off a meeting where the theme was national and transnational security implications of Big Data in the life sciences. We really wanted to tackle some of the security implications in the area of Big Data, where biology has almost a complete overlap with the digital world. At this meeting I had representatives from Microsoft, Intel, IBM, Google, and Amazon, the entities leveraging this Big Data bio-innovation future, and challenged them at the outset to identify potential security issues. We did find some significant issues and published some reports that are now publically available.

MBoC: You had an incredible lineup of expertise contributing to the AAAS report *National and Transnational Security Implications of Big Data in the Life Sciences* (Berger and Roderick, 2014). Was there any specific event that motivated the FBI to launch this reflection on biosecurity or was this entirely a proactive endeavor?

SSA You: The anthrax mailing in 2001 was a huge seminal event. Security discussions in the past tried to overlay security structures that were used in the nuclear or chemical realm. Completely locking down certain areas of expertise or materiel is completely antithetical to how the life sciences operate. If our mission is one of preventing the misuse, exploitation, or abuse of the life sciences, how do we approach security without becoming a hindrance to the life science enterprise?

Over the last two years, we have had the issues with regard to the Centers for Disease Control and Prevention (CDC) and Department of Defense (DoD). A lot of discussion also came when the J. Craig Venter Institute synthesized that bacterial genome. There were a lot of calls and discussions about the scientific community needing more ethics training and the need to develop a greater culture of responsibility. From a law enforcement perspective those are necessary but not sufficient. What has been lacking is the scientific community being provided security awareness—something that augments how they approach the life sciences. Individuals, no matter where they are in the world or when they enter the life sciences, always start with the premise, "Do No Harm," taking a page from the Hippocratic Oath. Unfortunately there are groups and individuals who do not subscribe to the same ethics and norms and agreements to integrity that we all take for granted and are almost innate for us. How do we graduate from "Do No Harm" to "Not On My Watch"? It means you take an active role in being sentinels for what you are doing and preventing the abuse, misuse, and exploitation of the life sciences. If *you* are not fully aware what the security vulnerabilities are, then that becomes a true vulnerability for all of us.

We also have a Biological Weapons Convention (BWC). It is amazing to me that we have an international treaty to which we are all beholden, yet there are very few programs, if any, in which incoming biology students are exposed to it or the fact that the BWC exists because biology had been absolutely used and exploited for

offensive purposes, even by the United States. If we do not teach that little bit of history and other security aspects, then it becomes really a challenge in the future on how to better protect biology. It is not about ethics; it is not about responsibility; it really is about having a healthy appreciation of some of the security concerns.

MBoC: How real is the threat? The aforementioned AAAS report read, "very little, if any, information exists about the theft, manipulation, or exploitation of Big Data in the life sciences."

SSA You: That is the key question. One of the goals of generating this report was to galvanize people to start thinking about security because quite honestly I do not think we really appreciate how deep or how wide the security vulnerabilities are in leveraging these large data sets or Big Data in general.

Referring to my prior comments about precision medicine, it all hinges on genetic information and longitudinal data over time. You are only as good as the size of the data set. You need a large data set because when you do an analysis you need statistical significance to know whether your results are right. As you think about that, let me walk you back to some of the most significant cyber-intrusions this past year. In August 2014, there was a Community Health Systems hack with 4.5 million patient records accessed; a few months after that was the large Anthem Blue Cross hack with 80 million individuals impacted; and then a month after that, the Premera Blue Cross hack in which 11 million patient records were hit. This is when it keyed off for me. In the Premera Blue Cross hack, clinical data were accessed too. Across the government, with these particular intrusions, the focus has been only on the potential loss of personal identifying information, the risk of fraud, and identity theft. I do not want to give that short shrift, because we are talking about tens, hundreds of millions of dollars in potential loss. However, if you think about the critical data—a beautiful longitudinal data set, containing an individual's demographics, disease state, drugs administered, and treatment received—someone now has a treasure trove of clinical trial information. Unfortunately, all of those hacks were allegedly attributed to a hacking group based out of China. It has become not just fraud anymore. There is a much broader security vulnerability, the potential loss of our ability to stay globally competitive in the new drug market. Now somebody out there has the brass ring—this gigantic data set, where the only limitation is deriving the analytical tools to make all that data useful.

There are a couple of issues now. One is to identify how much we have given up. We have to get beyond the paradigm of just looking at the financial loss. In the area of Big Data with specific applications to the life sciences, information taken could potentially be used for exploitation or extortion. A second is that, with the analytical tools that are coming online today, it will be almost impossible to deidentify information in the future. This was a key takeaway from the meeting with the AAAS-UNICRI last year. If you have any short genetic sequence of an individual, you can effectively deanonomize it in fewer than three steps with publically available tools.

MBoC: Privacy does not exist anymore?

SSA You: Correct. If you are part of an institutional review board, you are in big trouble in maintaining compliance and keeping up with protecting human subject information. That is just one regulatory hurdle that will be coming up.

MBoC: Where is the greatest security threat to Big Data through hacking? Is it through the lone wolf, companies engaged in industrial espionage, or is it from state-sponsored activities?

SSA You: My answer is "yes." The vulnerabilities are across the spectrum.

MBoC: Are the threats to Big Data greater for private Big Data, for example Pharma, or for academic Big Data?

SSA You: It is all of the above. Our AAAS meeting came to the crux of it: whoever has the largest and most diverse data set is going to win. That means we really need to start thinking in a more holistic manner what security means with a data set.

MBoC: Does the FBI define Big Data in terms of volumes of data or analytical functions? Is the threat against the volumes of data or the ability to analyze data?

SSA You: It is both. I do not want to go too far into definitions because one of the issues is how to define Big Data.

From a life sciences standpoint, we need to be going into this with our eyes wide open. How do we do anything? A thorough assessment of potential security vulnerabilities is a first step. Second, identify how to mitigate them up front. Finally, ask whether we have to come up with novel ways to address security in this bio-future. The power of the life sciences is open source, open sharing, but in it there is the added dimension of an individual's very intimate information. So there may be a call to redefine how we address security in the future. It may not be building up secure walls, whether they are physical or virtual, that protect data like our financial data. In this world of the life sciences, which is inherently open, we are going to have to rethink security.

MBoC: How should life scientists, faculty members at universities, respond to the worries of the FBI in terms of biosecurity? What do you see people doing to improve the situation?

SSA You: To me, the strategy is that once we build trusted partnerships with the scientific community, first with the FBI reaching out and providing the security awareness and education, something really profound happens. We have seen it happen in synthetic biology. You see the scientific community doing their own assessments of their technologies, self-identifying potential security vulnerabilities and then providing notification to the FBI—to my unit or other partners at the FBI. So the tables have turned. The scientific community educates the FBI on emerging vulnerabilities. They do us a favor, helping us to be better informed to better protect the life sciences, universities, and communities. Even better, the community will then develop security solutions based on their expertise, which is the best of both worlds. How powerful would that be when the experts, who are developing these powerful tools and applications of the future, immediately, on the front end, start developing and implementing security measures within these applications? That is where we want to be; that is where the future has got to be. So there is absolutely a very necessary and important partnership between law enforcement and the scientific community. It is just not a one-way street.

Take, for example, the scientific papers regarding CRISPR/Cas9 and gene drives and most recently the genetically modified yeast producing opioids. Scientists drafted the scientific manuscript and a companion editorial piece calling out the potential security vulnerabilities. That is powerful; that is a home run. We have successfully empowered the scientific community to understand security and then to take some proactive actions of their own.

MBoC: It seems one of the concerns of your unit, the BCU, is dual use of data. Does the BCU have formal relationships or work with the National Institutes of Health (NIH) National Science Advisory Board for Biosecurity or the CDC?

SSA You: Thank you for that question. It goes to the background of the WMD Directorate. One of the cornerstone aspects of our program is the really important position called the WMD Coordinator. These are men and women, Special Agents, trained in chemical, biological, radiological, and nuclear matters. We have at least one stationed in each of our 56 field offices across the United States. The WMD Coordinator's role, as the name implies, is to coordinate and

lead the notification protocols with state and local law enforcement, public health, partner with other federal entities, and then build relationships with universities, companies, and institutions within their jurisdiction. So if anything did occur, a local university, for example, would then know they have a local federal representative that can respond. If there is ever a biological incident or actual bio-crime, then those WMD Coordinators become critical in the response. They actually have been a big part of the action over the years with the DoD and CDC events, the discovery of smallpox at NIH at a Food and Drug Administration laboratory, the two high-profile ricin mailings almost two years ago, and the incident at Georgetown University where a student was manufacturing ricin in his dorm room. In all of these different incidents, those WMD Coordinators were called in and were part of the response. No matter where in the government, the Coordinator is there to help and assist with either preventive training or, if anything did occur, the response.

MBoC: How do people find the WMD Coordinators should they ever need one?

SSA You: They can just call the FBI field office in their jurisdiction and ask to be referred to the WMD Coordinator. Should any suspicious or criminal activity be observed that puts personnel, institutions, or materials at risk, contact your local FBI WMD Coordinator to help with any assessments. Think of them as being a resource to the scientific community. If you call them, it is not immediately the opening of an investigation. They are someone specifically within the FBI who is familiar with the life sciences community and with whom you can just touch base to see if something passes the sniff test.

MBoC: Although many readers of *Molecular Biology of the Cell* are gaining a greater awareness of Big Data, their own research does not take them into the realm of Big Data. For those readers, does the FBI have biosecurity concerns that lie with small data or is the focus really on Big Data?

SSA You: The focus on Big Data is because it is an emerging area. If you see all of our activities, the overall theme is safeguarding science, whether you are working in large data analytics, with select agents, yeast, or *Escherichia coli*. We are not honing in on a specific subgroup or subtopic of the life sciences. It is really preventing the misuse of the life sciences in general.

MBoC: The FBI's primary role is safeguarding the homeland, the United States. Many of our readers are not Americans. Is there a separate, special, or additional message for people doing life science research, especially Big Data research, outside the boundaries of the United States?

SSA You: Safeguarding science is universally applicable. I hope for a future when biologists are working as WMD Coordinators in other law enforcement agencies around the world. We need that. The 21st century will see the same leaps and bounds with the life sciences that we saw in the 20th century with the Internet and personal computing. If there is going be a global impact from the life sciences, there is absolutely a call to action for biologists wherever they are in the world to be guardians of science. However, we need to come to a realization first that there will be issues. We have to start discussing these things now before it is too late, before any attempts at security will be too little, too late. We are much better served tackling issues sooner. I hate to say it, but, if we are not careful and there is a complete overlap of the life sciences and the digital world, we might see ourselves with our security as we are facing cyber-security right now, and we do not want to be in that position.

FROM BENCH TO BADGE—ARE YOU HIRING?

MBoC: It is clear there is a lot of work ahead, not just for the scientific community, but for the FBI as well. What are the career opportunities for cell biologists in the FBI, whether they have Big Data experience or not?

SSA You: We are most definitely hiring. You can be a Special Agent like me, or there are support positions such as the scientists who work in our laboratory division. These are individuals who develop the tools for forensic analysis. A key piece of our mission is looking at intelligence; that is an analyst position. There will absolutely be a need for folks with a biology background. You do not need to have law enforcement experience. I did not.

I will be completely candid, upfront—our hiring is a very competitive process. Prior to 9/11, the FBI's focus was on hiring individuals with law enforcement or military experience, lawyers, or accountants because the primary mission was tackling organized crime. In this day, when our number one priority is prevention, there is an absolute critical need for hiring individuals with background in computer science, foreign languages, and especially the natural sciences. If you have a chemistry or biology background, you are in the running. The minimal criteria are a bachelor degree and at least three years of real-world experience. More than anything else, if you can articulate and show how you excelled in your specific field, then you are a good candidate. Your field does not necessarily have to be Big Data. You need to be passionate about what you do because in doing so, you inherently excel. The key is to set yourself in a position where you can really excel so when we begin talking to your coworkers and managers about who you really are, you have put them in a position where they can say you are an integral part of the team and made significant contributions. That will be a good selling point for a future career in the FBI.

THE TAKE-AWAY

SSA You: Partnerships between the FBI and the scientific community to build security awareness are essential. Big Data in the life sciences is taking the biosecurity discussion beyond pathogens and toxins. Historically, the conversation almost always fell on pathogens and almost exclusively on select agents. We have to widen the aperture of what we mean by biosecurity in the future.

ACKNOWLEDGMENTS

I thank John Fleischman, American Society for Cell Biology Senior Science Writer, for tips in the preparation of this interview.

REFERENCE

Berger KM, Roderick J (2014). National and transnational security implications of Big Data in the life sciences. Available at www.aaas.org/report/national-and-transnational-security-implications-big-data-life-sciences (accessed 19 March 2015).